The Predictive Aspect of Business Process Intelligence

Lessons Learned on Bridging IT and Business

Pérez, Moisés Lima; Møller, Charles

Published in:
Business Process Management Workshops

DOI (link to publication from Publisher):
10.1007/978-3-540-78238-4_3

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

The Predictive Aspect of Business Process Intelligence: Lessons learned on bridging IT and business

Moisés Lima Pérez[1], Charles Møller[2]

[1] Microsoft Denmark, Tuborg Boulevard 12, 2900 Hellerup, Denmark, moisesp@microsoft.com
[2] Aarhus School of Business, University of Aarhus, Fuglesangs Allé 4, 8210 Aarhus V, Denmark, charles@asb.dk

Abstract. This paper presents the arguments for a research proposal on predicting business events in a Business Process Intelligence (BPI) context. The paper argues that BPI holds a potential for leveraging enterprise benefits by supporting real-time processes. However, based on the experiences from past business intelligence projects the paper argues that it is necessary to establish a new methodology to mine and extract the intelligence on the business level which is different from that, which will improve a business process in an enterprise. In conclusion the paper proposes a new research project aimed at developing the new methodology in an Enterprise Information Systems context.

Keywords: Business Process Intelligence; Data Mining, Enterprise Information System; Customer Relationship Management.

1 Introduction

In order to stay competitive in dynamic environments, companies must continually improve their processes and consequently align their business, people and technologies. Some companies have built their businesses an their ability to collect, analyze and act on data [1]. The ability to accurately predict consumer demand coupled with the capability to rapidly react and readjust to environmental changes and customer demand fluctuations separates the winners from the losers [2].

The emergent challenge and also opportunity for an organization is to be able to design and to orchestrate adaptive global supply chain networks. Adaptive supply chain networks exploit process innovations to improve efficiency and responsiveness and consistently achieve these objectives. Furthermore most global supply chains are vulnerable to disruptions like natural disasters, accidents and terrorism. Consequently not only sense and response capabilities are relevant [2], but also robustness and resilience is an emerging concern to most enterprises [3].

Agility in a global context is inevitably tied into technology and modern Enterprise Information Systems (EIS) from the major vendors such as SAP, Oracle and Microsoft include the concepts and tools needed for creating a flexible infrastructure [4]. However, there is still a gap between the flexibility the technology provides [5] and the agility needed by companies, and this gap is expected to exist for a while [6].

However we see an increasing research focus on process driven information systems [7] and the issues of relating data to the process model [8].

This paper suggests that there is a huge potential contribution in using advanced EIS to transform an entire supply chain and create a better alignment between business and IT. The management of business process and thus the concept of Business Process Management (BPM) are central and one of the techniques is process intelligence (BPI). The ability to predict is one of the important features BPI will have to offer to achieve the aim of improving the business processes. The importance for BPI to predict events has already been highlighted in previous studies [9]. In this paper we will explore the predictive aspects of BPI and based on an analysis of a case study we call for a new approach to BPI that addresses the integration of technology and management. First we will present and discuss the existing research on BPI in the next chapter in order to identify the gap.

The background for the case on prediction is based on a real data mining project, where a trade union institution in Denmark needed to predict churn among its members. After the case study the article discusses the need to look beyond process log information to enable users to conduct effective prediction. Based on the case it is concluded that besides these logs you often need traditional BI information such as customers' demographics to supplement process models. Consequently the paper proposes a research project aimed at developing a new methodology for BPI.

2 Business Process Intelligence

Business Process Management (BPM) may still be an immature concept [10] but BPI has yet to be established as a concept. The concept is used by a group of HP researchers to capture a set of tools in the BMPS suite [11, 12], but is there more in the concept?

Casati et al. explicit states: "we assume that it is up to the (business or IT) users to define what quality means to them, and in general which are the characteristics that they want to analyze" [11]. Grigori et al. focus on a set of tools that can help business IT and users manage process execution quality [12]. They call this set of tools Business Process Intelligence, where they apply business intelligence techniques to business process. They extract, cleanse and integrate data from BPM systems (only process logs) and aggregate them in a data warehouse. From this information reservoir they can then extract information on high or low performance of business processes. They propose a BPI suite that is made up of these modules:

- Analysis module: it allow users to analyze the execution of a process
- Predict: to find outliers (processes that under perform)
- Monitor: send alerts when these outliers are identified
- Control: mitigate or reduce the consequences of a negative outlier
- Optimize: improve a business process based on prediction data

They acknowledge the fact that to achieve these aims, a number of hurdles have to be overcome such as the choice of architecture and technology, strategy to process the information, the choice of tools that make it easy for users to understand the outcome of this information, etc…

Their research, called Business System Intelligence, is aimed at developing techniques that allow analysis, prediction, monitoring, control and optimization of business processes. In their paper they explain the concepts used to process the logs, the architecture and semantics used in their data warehouse that stores this information and the analytics and prediction models used in their cockpit [13].

They plan to work on challenges as duplicate data and extend the mining capabilities to text mining as well as the use of decision trees.

Recently we have also seen the emergence of the Business process mining concept [14]. Business process mining takes information from systems as CRM and ERP and extracts knowledge from them that can then be used to improve a given aspect of a business. The paper contends that few of the software vendors that can provide these features are tested in real life. The paper describes the process mining project they conducted in Holland to analyze contractors' invoices from three perspectives: case, organizational and process perspectives. The organizational perspective focuses on the performers of a task and how they are related. The case perspective focuses on the properties of a case or task and the process on the order of activities.

The mining models developed revealed undesired loops in the work process. These loops have a great impact on the performance as whole. They conclude that process mining is a reality but that one needs to contend with noise in the data.

However there is more to a business process than just the workflow [15]. In the paper the authors highlight the intrinsic gap between business models (as business people understand their work) and work flow specs (as IT people describe business processes). The first one is the result of an intuitive process while the second one aims at building process-aware systems.

The paper proposes a five step method to guide IT modelers in building workflows. They use Petri nets as the graphical technique. The five steps are business process modeling, transformation into work flow nets, correctness check, strategy determination and strategy implementation. They raise the awareness that understanding a business is paramount for any interpretation or improvement of a process. A formal language is not enough since there is an intrinsic gap between the business as is and its interpreter. A methodology is required to cover this gap.

ERP systems are used to support business processes and there is a current need to verify that these systems are well-tuned and perform effectively. With this in mind Ingvaldsen & Gulla [16] presents a case that used ERP logs data to build mining models that could be used to improve business processes in Norway. The mining tool was built for a large agricultural dealer.

The tool has 3 modules: a model (that describes the logic of a business process), statistical analysis of key performance indicators and knowledge discovery that combines data from other sources like data warehouses with process logs. In consistency with previous studies (e.g. [12, 17]) they also saw the need to combine data from external sources, such as the department and employee involved in a process with actual process logs to achieve better knowledge discovery results. The results of this case were hampered though by lack of sufficient data, which indicates they might have missed important trends and might have highlighted as trends processes that were actually outliers.

There is evidence that points towards the fact that they may not have obtained the right information on processes. List & Machaczek [18] highlights that one major shortcoming of current performance measurement is that business processes are not measured systematically. Their aim is to integrate process performance measurement into the corporate data warehouse. For this, they conducted a case study in a large insurance company. They propose a performance measurement system that collects process related data and links it to a corporate data warehouse. The purpose is to obtain a holistic view of the corporate performance. It was important to do this information gathering constantly to avoid the time lag common to other measurement systems. The case shows the potentials that lie in using traditional methods of data warehousing to process and extract knowledge from process logs. They use both traditional data models (star schemas) and olap tools for their analytical platform.

But also traditional BI projects are facing issues. Missi el al. [19] were looking at existing methods and conducted studies on several BI projects and conclude that lack of proper information integration can be pointed as one major cause of failure. By reviewing existing literature about data integration they attempt to provide a framework for effective data processing in BI systems. The research used CRM systems as focus. They have created a taxonomy for data quality problems (e.g standardization and matching), and integration problems. They then go on a detailed analysis of existing ETL tools and their weakness to overcome the problems inherent in data processing. Clean and integrated data seems to be the only way towards meaningful BI.

In conclusion a recent study by Davenport [1] shows the importance of predictive analytics in any kind of activity from business to sports. It highlights those companies and even sports teams that use statistics to improve their processes and win competition. But it also contends that you need experience to correctly interpret the numbers you get from statistics and take the right decision.

Davenport contends that there are a number of factors you need to add to the statistics to make the best use of it:

- The right focus: choose those business areas that make a difference for your organization (customer service for instance)
- The right culture: which can go from the very intuitive approach to a rigorous measurement of a product performance
- The right people: winning organizations make sure to hire the right people to do their analytics
- The right technology: some companies go as far as building their own super computers because existing technology cannot cope with their demands

The article highlights the fact that even when you cannot compete with a rival due to price, you can still beat them on process effectiveness. These winning companies attribute their success to their ability to exploit data and to extract intelligence from it which constantly improves their processes.

In the next section we will evaluate the experiences from a business process improvement project which took a traditional BI perspective.

3 Case Study: BI in a Danish Trade union

The case company is one of Denmark' largest trade unions but it is annonymized in the study. The company have in recent years faced problems with customer loyalty. Their essential problem is that their churn rate (10 %) has been higher the rate of customer acquisition (2 %). Last year they were interested in learning what data warehousing and data mining could do to explain the reasons for the customer churn.

The trade union in question was already using CRM so they were interested in checking the efficiency of their existing services. First, a workshop was conducted in order to understand the nature of those that churned. Before deciding on any particular algorithm of data mining, it was decided to follow the steps elaborated in the proceeding sections.

3.1 Definition of a business issue to predict: Success criteria

It was of paramount importance was to conduct a business analysis session, where we focussed on understanding the types of services they provided, which segments of the population they worked with, and in their opinion, the reason of the problem they were facing.

All this background information was needed to determine the scope of the workshop. We also needed to determine the type of prediction they wanted and agreed on building a model that identified those customers churned against those that did not.

Although there were a number of different reasons for churning (change of work, economic reasons, retirement, deaths, etc.) we decided to group them in a binary form, that is distinguish those that churned (marked as 0) against the loyal ones (marked as 1).

The success criteria for the prediction model were decided to be at least 60 percent of accuracy for both those that churned and those that did not.

The aim was to improve their customer handling process in their CRM system; so that they could improve their services over the phone (e.g. give special attention to a member who was likely to churn). They intended to use this information also to improve on the legal and educational offers they offered, especially to those that were possible churners.

3.2 Data Analysis and Preparation

The next step was an initial analysis of available data and its preparation. Data analysis was to yield two important results: the quality and accuracy of the data and its relevance to the business aim at hand, namely churn prediction.

Quality problems as duplication, incomplete information, values out of range, missing values etc. are known to have a negative bias on a data mining exercise. Therefore we determined a number of quality tests that the data had to pass.

Accuracy of the information was also of great importance and therefore we needed an experienced person in the trade union that could help us set up the logic that the data was supposed to comply with. These rules were also used to select the data and transform it when needed.

For the exercise we used SQL server 2005 data mining suite. This suite is able to calculate the accuracy of the information in relationship to the issue to be predicted. So during this early stage of the data mining exercise we were able to spot those attributes of almost no relevance and exclude them from the exercise.

One important discovery, which is of relevance to BPI, was that none of the information which came from their CRM systems, such as complaints, types of transactions the customers made etc. were of relevance to determine the likelihood of whether a customer would churn or not.

We therefore concentrated on data that was stored in their legacy system such the demography of their customers, the number of years they had with the union before they left it, the type of education they had, the work they performed, etc.. We also insisted on collecting as much information as possible on both those that churned and those that did.

This step was the one that consumed the greatest effort, since we had to resolve all data quality issues first in order to avoid biasing the models with incorrect information.

3.3 Selection of training and validation sets

To ensure the effectiveness of the exercise we divided the data into two sets: one set was to be used to train the model and the other one to test the model. It is important that this division is done by some sort of random selection, so that you would avoid bias in either the training or the test set.

One of the issues that caused discussion was the percentage of data that should belong to either the churn or the non-churn side in the training set. One argument was to use the same proportion as in the real world. In this case 10 percent should have been the churner and the rest the none-churners. However, it soon showed that the models based on this proportion were not effective, that is the level of the prediction accuracy was way below the expected (only 45 %). After several trials, the ideal proportion for the training set was 50/50. This proportion proved to help our models to “discover” the patterns behind each group and effectively predict real life situations. The test set had to reflect reality so that we were sure to later use it in real life. So we built a test set that contained only 10 percent of churners and 90 percent of loyal members.

3.4 Data Mining Models: Development

Data mining tools should provide the possibility of comparing models with the same training set. The transition from one algorithm to another should be an easy one, since in this phase you will need to test different strategies, re-evaluate the source data you use, probably change the values of some attributes (from continuous values to bindings in ranges, etc.).

Data mining development also implies that you need to compare the effectiveness of the models used. SQL 2005 data mining suite comes with several algorithms such as e.g. Decision Trees, Naïve Bayes, Clustering, Neural network, Sequencing and

Association. For our case, we used decision trees, Naïve Bayes and Clustering. We found that models based on decision trees were the most effective to predict a customer's likelihood to churn or not.

All our decision tress identified that membership fee was the most relevant factor. This came as quite a surprise to the trade union. They had worked hard to keep membership fee as low as possible. They had even conducted a number of statistical studies internally where they had ruled out membership fee as a determining factor for customer loyalty. In our case, seniority and work trade came as the next most relevant factors. Again none of these attributes were used in any of their CRM processes and therefore we needed to look for churn reasons in their legacy systems instead of their process logs.

3.5 Model Validation

The platform we used allowed us to validate the model against both a so-called perfect model (the green line) and another one called random model (the blue line) as illustrated in figure 1. In data mining language this is called to "assess the model's lift" or its degree of accuracy. A proof of this models accuracy is the fact that it stays close to the perfect model. This diagram also shows that the decision tree model will perform perfect with up to 30 percent of the population when it predicts those customers that churn. After that the performance degrades somewhat and starts picking up again when it reaches 80 %.

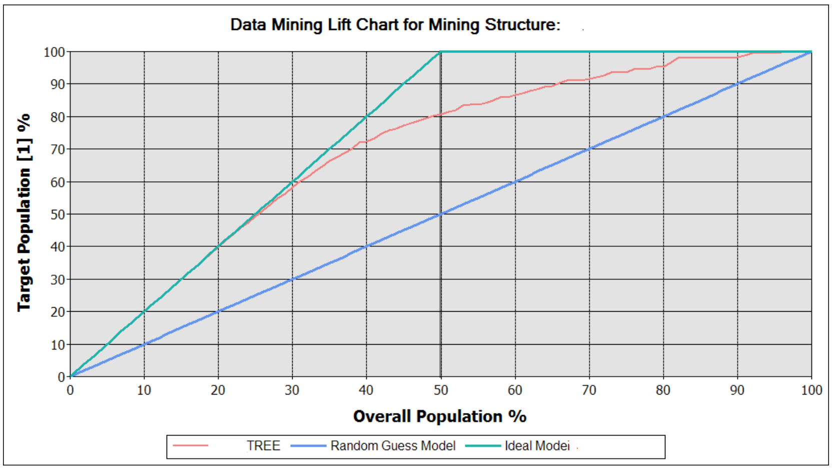

Figure 1. Model validation

We also wanted to test models against each other. This allowed us to see which model to use for future tests. The following picture shows in different colours the models we developed. The green one again is the so-called perfect model. The next one in effectiveness is the decision tree model. The Bayes and the cluster model did not perform as well as the decision tree.

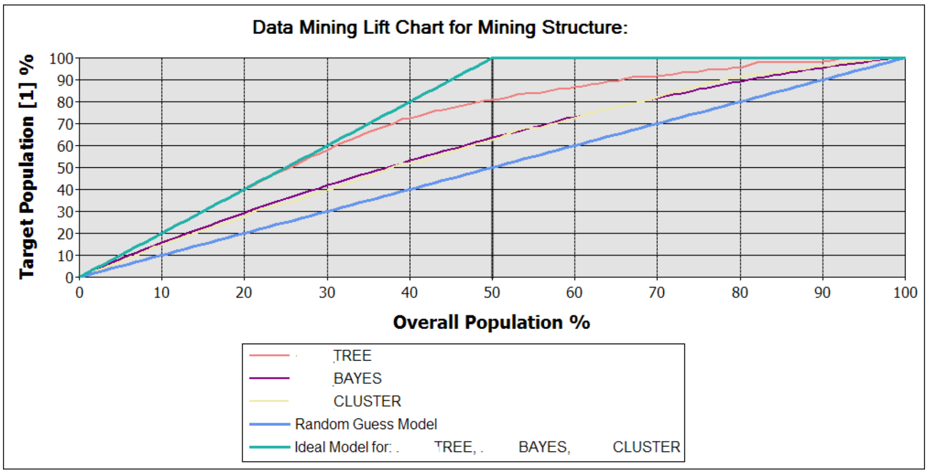

Figure 2. Model comparison

3.6 Model test with completely new data

The tests were conducted with a completely new set of data. In our cleansing process we ensured to use data that was uniquely identified by the member id. All duplicate records were eliminated and then we had a large set that was randomly" divided into two sets. The training set had 90 percent of its records as loyal customers and 10 as churning. The proportion between the two sets reflected the trade union's actual loyal vs. churn rate.

We chose the decision tree model since it was the one that showed best performance (accuracy to predict churn and not churn records).

Except for the churn attribute, all the other fields from the test set (such as membership fee, education, work type, seniority in the union) were entered in the test mode and the model should then predict the churn value. Predicted churn versus real churn is presented in table 1. The value 1 represents those that were loyal members and the 0 those that churned. The green row illustrates where the model had a 92 percent of accuracy for the loyal cases and a 74 percent for the churn-cases. The "Bar of Excellency" decided at the beginning of the workshop was of 65 percent for either case (the total amount of loyal ones were 90000 and 10000 those that churned).

Table 1. Predicted and real churn

Predict	Real	Count
0	0	82791
0	1	2628
1	0	7209
1	1	7372

4 Lessons learned and discussion

From a Data mining perspective the trade union case is trivial but with the developed model they are able to enhance their customer service processes considerably. In the case study we discovered that the most business relevant information was found in the legacy enterprise systems and not in the process logs.

Most information mining projects fail due to lack of a proper method. One of the key issues in such an exercise is to start with a clear business goal which should be quantifiable. It is also crucial in any data mining methodology to find relevant and cleansed information from which to develop a model.

This implies that BPI should be considered on two levels: 1) on the system or the BPMS level where most of the present research has been focusing; and 2) on the business level where the contextual information and business issues direct the prediction effort. These two approaches are quite different but we are suggesting that they can supplement each other.

Where BPI on the system level with existing technologies can be automated through formal models, BPI on the business level is directed by “ad-hoc” approaches on a one-shot basis. In order to have business level BPI we need to develop approaches to business process and service modelling in the business domain, and then to apply the techniques from system level BPI.

Consequently we advocate that BPI research should enhance its perspective from process logs towards a “holistic” approach where process-derived data is merged with general enterprise system information. Pre-processing this enterprise information through a BI strategy will give a better picture of what elements a BPI model should include and substantially reduce the time needed in the processing efforts to identify the best predictive model for business impact.

5 Conclusion

In this article we have argued that BPI is important to a modern global enterprise and we have emphasized prediction as a key characteristic of the business value of BPI. Through the case study we have argued that existing BPI techniques are missing an important business link and that this link can be extracted from existing enterprise systems. Finally we concluded that research needs to extend its perspective towards these business issues.

This leads to the formulation of a new research project on Business Process Intelligence in the context of a global business [20]. One of the research challenges in this project is to transfer the methods and techniques from the systems level BPI towards business level BPI. The long term vision is to automate business improvement activities using BPI.

Initial studies suggest that a global business with a large and complex enterprise systems setup obtains a substantial benefit from this approach.

References

1. Davenport, T.H.: Competing on analytics. Harvard Business Review **84** (2006) 98-107
2. Rai, A., Ruppel, C., Lewis, M.: Sense and Respond. University Thought Ledership forum. Georgia State University, Rollins College, Atlanta, Georgia (2005) 17
3. Sheffi, Y.: Building a Resilient Supply Chain. Harvard Business Review Supply Chain Strategy **1** (2005)
4. Møller, C.: ERP II: a conceptual framework for next-generation enterprise systems? Journal of Enterprise Information Management **18** (2005) 483-497
5. Fingar, P., Bellini, J.: The Real-Time Enterprise: Competing on Time with the Revolutionary Business S-Ex Machine. Meghan Kiffer Press (2004)
6. Melenovsky, M.J., Sinur, J.: BPM Maturity Model Identifies Six Phases for Successful BPM Adoption. Gartner Research, Stamford, CT (2006) 14
7. Dumas, M., Aalst, W.M.P.v.d., Hofstede, A.H.M.t.: Process-Aware Information Systems: Bridging People and Software through Process Technology. John Wiley & Sons Inc. (2005)
8. W.M.P. van der Aalst, H.A.R., A.J.M.M. Wejters, B.F. van Dongen, A.K. Alves de Medeiros, M. Song, H.M.W. Verbeek: Business Process Mining: An Industrial Application ELSEVIER (2006) 20
9. Grigori, D., Lorraine, L.I., Casati, F., Dayal, U., Shan, M.-c.: Improving Business Process Quality through Exception Understanding, Prediction, and Prevention. Proceedings of VLDB'01, Rome, Italy (2001)
10. Møller, C., Tan, R., Maack, C.J.: What is Business Process Management? A two stage literature review of an emerging field. CONFENIS 2007 - The IFIP International Conference on Research and Practical Issues of Enterprise Information Systems, Beijing, China (2007)
11. Casati, F., Dayal, U., Sayal, M., Shan, M.-c.: Business Process Intelligence. (2002)
12. Grigori, D., Casati, F., Castellanos, M., Dayal, U., et al.: Business Process Intelligence. Computers in Industry **53** (2004) 321-343
13. Sayal, M., Casati, F., Dayal, U., Shan, M.-C.: Business Process Cockpit. Proceedings of the 28th VLDB Conference, Hong Kong, China (2002)
14. van der Aalst, W.M.P., Reijers, H.A., Weijters, A.J.M.M., van Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W.: Business process mining: An industrial application. Information Systems **32** (2007) 713-732
15. Dehnert, J., Van Der Aalst, W.M.P.: Bridging the gap between Business Models and Workflow Specifications. International Journal of Cooperative Information Systems **13** (2004) 289-333
16. Ingvaldsen, J.E., Gulla, J.A.: Model-Based Business Process Mining. Information Systems Management **23** (2006) 19-31
17. Muehlen, M.z.: Process-driven Management Information Systems — Combining Data Warehouses and Workflow Technology. Proceedings of the International Conference on Electronic Commerce Research (ICECR-4),. IEEE Computer Society Press, Los Alamitos, California,, Dallas, TX (2001) 550-566

18. List, B., Machaczek, K.: Towards a Corporate Performance Measurement System. SAC'04, Nicosia, Cyprus (2004)
19. Missi, F., Alshawi, S., Fitzgerald, G.: The Impact of Information Integration on the success of Business Intelligence Projects. AIM 2003, Grenoble, France (2003)
20. Møller, C.: The Conceptual Framework for Business Process Innovation: Towards a Research Program for Global Supply Chain Intelligence. The Icfai Journal of Supply Chain Management **Forthcoming** (2007)